EpiBuffer: Modeling Spatial Heterogeneity in Exposure Buffers and Risk

SpatialBuffers Statistical Model

$$Y_i\left(\mathbf{s}_j\right)|\mu_i\left(\mathbf{s}_j\right), \boldsymbol{\zeta} \overset{\text{ind}}{\sim} f\left(y \mid \mu_i\left(\mathbf{s}_j\right), \boldsymbol{\zeta}\right),\ j = 1, ..., m,\ i = 1, ..., n_j;$$
$$g\{\mu_i(\mathbf{s}_j)\} = \mathrm{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\}\theta\left(\mathbf{s}_j\right).$$

Likelihood Options

- Binomial likelihood with logit link function (likelihood_indicator = 0):

$$Y_i(\mathbf{s}_j)|p_i(\mathbf{s}_j) \overset{\text{ind}}{\sim} \text{Binomial}\left(\tilde{n}_i(\mathbf{s}_j), p_i(\mathbf{s}_j)\right), \text{logit}\left\{p_i(\mathbf{s}_j)\right\} = \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\}\theta\left(\mathbf{s}_j\right);$$

- Gaussian likelihood with identity link function (likelihood_indicator = 1):

$$Y_i(\mathbf{s}_j) = \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\}\theta\left(\mathbf{s}_j\right) + \epsilon_i,\ \epsilon_i|\sigma^2_\epsilon \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma^2_\epsilon\right);$$

- Negative binomial likelihood with logit link function (likelihood_indicator = 2):

$$Y_i(\mathbf{s}_j)|r, p_i(\mathbf{s}_j) \overset{\text{ind}}{\sim} \text{Negative Binomial}\left(r, p_i(\mathbf{s}_j)\right), \text{logit}\left\{p_i(\mathbf{s}_j)\right\} = \mathrm{O}_i(\mathbf{s}_j) + \mathbf{x}_i(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\}\theta\left(\mathbf{s}_j\right).$$

Exposure Definitions

- Counts (exposure_definition_indicator = 0):

$$\mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\} = \sum_{k=1}^{h} 1(\mathrm{d}_{jk} \le \delta\left(\mathbf{s}_j\right));$$

- Spherical (exposure_definition_indicator = 1):

$$\mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\} = \sum_{k=1}^{h} 1(\mathrm{d}_{jk} \le \delta\left(\mathbf{s}_j\right))\left(1 - 1.5\left\{\frac{\mathrm{d}_{jk}}{\delta\left(\mathbf{s}_j\right)}\right\} + 0.5\left\{\frac{\mathrm{d}_{jk}}{\delta\left(\mathbf{s}_j\right)}\right\}^3\right);$$

- Presence/absence (exposure_definition_indicator = 2):

$$\mathrm{z}\left\{\mathbf{s}_j; \delta\left(\mathbf{s}_j\right)\right\} = \min\left(\sum_{k=1}^{h} 1\left\{d_{jk} \le \delta\left(\mathbf{s}_j\right)\right\}, 1\right);$$

  - d_{jk}: Distance between location $\mathbf{s}_j$ and point source $\mathbf{c}_k$;
  - h: Total number of point sources.

Spatially-varying Radii and Exposure Effect Parameters

- Radii:

$$\Phi^{-1}\left(\frac{\delta\left(\mathbf{s}_j\right) - a}{b - a}\right) = \mathbf{w}(\mathbf{s}_j)^{\mathrm{T}}\boldsymbol{\gamma} + \phi(\mathbf{s}_j);$$

  - $\boldsymbol{\phi}^{\mathrm{T}} = \left\{\phi\left(\mathbf{s}_1\right), \ldots, \phi\left(\mathbf{s}_m\right)\right\}|\tau^2_\phi, \rho_\phi \sim \text{MVN}\left(\mathbf{0}_m, \tau^2_\phi\Sigma\left(\rho_\phi\right)\right)$, $\Sigma_{ij}\left(\rho_\phi\right) = \exp\left\{-\rho_\phi||\mathbf{s}_i - \mathbf{s}_j||\right\}$;
  - Gaussian predictive process option also available for large m;

- Effects:

$$\theta\left(\mathbf{s}_j\right) = \mathbf{q}\left(\mathbf{s}_j\right)^{\mathrm{T}}\boldsymbol{\eta}.$$

Prior Information

$\beta_j \stackrel{\text{iid}}{\sim} \text{N}\left(\mathbf{0}, \sigma^2_\beta\right),\ j = 0, \ldots, p_x - 1;$

- p_x: Length of $\mathbf{x}_i$ vector (same for all i), including an intercept;
- Default setting: $\sigma^2_\beta = 100^2$;

$\gamma_j | \tau^2_\phi \stackrel{\text{iid}}{\sim} \text{N}\left(\mathbf{0}, \frac{1-\tau^2_\phi}{p_w}\right),\ j = 0, \ldots, p_w - 1;$

- p_w: Length of $\mathbf{w}\left(\mathbf{s}_j\right)$ vector (same for all j), including an intercept;

$\eta_j \sim \text{N}\left(0, \sigma^2_\eta\right),\ j = 0, \ldots, p_q - 1;$

- p_q: Length of $\mathbf{q}\left(\mathbf{s}_j\right)$ vector (same for all j), including an intercept;
- Default setting: $\sigma^2_\eta = 100^2$;

$\tau_\phi \sim \text{Uniform}\left(0, 1\right);$

$\rho_\phi \sim \text{Gamma}\left(a_{\rho_\phi}, b_{\rho_\phi}\right);$

- Default setting: $a_{\rho_\phi} = b_{\rho_\phi} = 1$;

Gaussian specific: $\sigma^2_\epsilon \sim \text{Inverse Gamma}\left(a_{\sigma^2_\epsilon}, b_{\sigma^2_\epsilon}\right);$

- Default setting: $a_{\sigma^2_\epsilon} = b_{\sigma^2_\epsilon} = 0.01$;

Negative binomial specific: $r \sim \text{Discrete Uniform}\left[a_r, b_r\right];$

- Default setting: $a_r = 1, b_r = 100$.

Default Initial Values

- $\beta_j = 0$ for all j;
- $\gamma_j = 0$ for all j;
- $\eta_j = 0$ for all j;
- $\tau_\phi = 0.50$;
- $\rho_\phi = -\ln\left(0.05\right) / \max\left\{||\mathbf{s}_i - \mathbf{s}_j||\right\}$;
- Gaussian specific: $\sigma^2_\epsilon = \text{variance}(\mathbf{Y})$;
- Negative binomial specific: $r = 100$.

Additional Information

- waic_info_indicator = 1: Provides output needed to calculate Watanabe-Akaike information criterion;
- fitted_info_indicator = 1: Provides posterior samples of fitted values;
- v: Required input; an $\sum_{j=1}^m n_j$-length vector with the location number that the observation is connected to (i.e., $1, \ldots, m$);
- exposure_dists: Matrix of dimension $m \times h$ (for h total exposure sources), specifying the pair-wise distances between all unique spatial locations (rows) and all exposure sources (columns). Units of exposure_dists should match the units of radius_range (e.g. both in kilometers)
- full_dists: Matrix of dimension $(m + k) \times (m + k)$, where the upper left m by m matrix describes the distances between the m observed locations, the bottom right k by k matrix describes the distances between the selected grid of $k << m$ locations (if the computational version is used), and the off-diagonal matrices describe the distances between the observed and grid locations (i.e., this full distances matrix is computed after stacking the observed locations and sampled grid locations into a single vector). If the

computational version is not needed, input the observed locations instead of the sampled grid locations when calculating these distances.

- It is recommended to scale the matrix of spatial distances (full_dists) by the largest observed distance in the matrix prior to running the model.

Interpretation

Within the function, exposure is scaled by the factor `exposure_scale`. As a result, the fitted θ and η values returned by the function depend on the exposure scaling factor used. The function returns the value used for scaling (`exposure_scale`), allowing the user to unscale as desired: $\theta_{unscaled} = \theta_{scaled}/\text{exposure\_scale}$; $\eta_{unscaled} = \eta_{scaled}/\text{exposure\_scale}$.